Design and Development of Medical Image Segmentation System for Early Diagnosis of Diseases

Synopsis Submitted to,

SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY

For Degree of Doctor of Philosophy

in

Computer Science

Under the Faculty of Science

By

Mr SHRIKANT SOMANNA

M.Sc. (Computer Science), M. Phil

Guide

Dr. V. T. Humbe

SRTMU Sub campas , Latur.

March 2023

Design and Development of Medical Image Segmentation System for Early Diagnosis of Diseases

Contents

1. Introduction 1
2. Research Challenges 3
3. Objectives 4
4. Literature Review 4
5. Proposed Work 11
6. Methodology 12
7. Result Analysis 13
8. Conclusion 13
9. References 15

1. Introduction

The process of segmenting a picture entails breaking it up into sections that are strongly correlated with the region of interest (RoI) in the original image [1, 2]. The purpose of medical image segmentation [3] is to provide a given input image in a meaningful form to study the anatomy, identify the region of interest (RoI), measure the volume of tissue to measure the size of the tumour, and assist in deciding the dose of medication, planning of treatment prior to applying radiation therapy, or calculating the radiation dose. The region of interest is highlighted via image segmentation, which aids in the study of medical images. MRI brain tumour boundary extraction, cancer detection in biopsy pictures, mass segmentation in mammography, border detection in coronary angiograms, segmentation of pneumonia affected area in chest X-rays, etc. are all examples of applications for segmentation techniques. Due to a lack of skilled labour, many medical picture segmentation algorithms have been created and are in demand [4]. The earliest image segmentation models were based on conventional approaches to image processing, such as thresholding and edge-based and region-based algorithms [3, 5]. In the thresholding technique, pixels were assigned to various categories based on the range of values within which they fall. The pixels in an image are classified as edged or nonedged according to the filter output when using the edge-based approach. In region-based segmentation techniques, groups of pixels with differing values and adjacent pixels with similar values were separated. Medical image segmentation is a challenging issue because of the limitations imposed by the method used to acquire medical images, the type of pathology, and distinct biological changes [6]. Medical imaging specialists are in scarce supply, yet they can analyse medical images [7]. Deep learning networks have helped to create better picture segmentation models with improved performance in recent years. On numerous well-known datasets, the deep neural networks had attained remarkable accuracy rates [8-10]. Semantic segmentation and instance segmentation are two main categories for picture segmentation approaches. Semantic segmentation can be viewed as a classification problem for pixels. Each pixel in the image is assigned to a certain class using this segmentation process [11-14]. Each object of interest in the input image is found and identified via instance segmentation. The recent literature in medical image segmentation is covered in the current work. The article reviews various deep learning-based picture segmentation algorithms and outlines their structural design. The task of reviewing medical picture segmentation has been worked on by numerous writers. In the realm of medical image segmentation, a few review publications that used deep CNN are described in Table 1 together with a public dataset in Table 2. The numerous deep neural

networks are covered in all of the aforementioned survey literatures. This review paper not only summarises the various deep learning methodologies, but also offers insight into the many medical picture datasets used for deep neural network training and clarifies the metrics used to assess a model's performance. The current work also highlights the numerous issues that DL-based image segmentation models have to deal with and their cutting-edge solutions.

Table 1 Comparison of deep learning methods in medical image segmentation

Organ	Segmented area	Model utilized	Dataset	Modality	Remarks
Cardiac	Cardiac, left, and right ventricular cavities and myocardium [15]	2D/3d CNN	ADC2017	Cardiac MR images	—
	Heart [16]	RFCN	MICCAI[2] 2009 challenge dataset	Cardiac MR images	RFCN reduces computational time, simplifies segmentation, and enables real time applications
	Heart [17]	U-Net	—	DT-CMR images	U-Net automated the DT-CMR postprocessing, supporting real time results
Brain	Brain tissues [18]	2D CNN	—	Multimodal MR images	Model performance increases by employing multiple modalities
	Brain tumor [19]	SDResU-Net	—	MR images	U-Net has generalization capability
	Brain [20]	Voxel-wise residual network	MRBrainS	MRI	—
	Brain [21]	DNN	ISBI 2012 EM	TEM	—
	Pixel-wise brain segmentation [22]	MD-LSTM	MRBrainS13	Brain MR images	It can parallelize for 3D data
	Brain tumor core [23]	FCN, U-Net		MR images	Bounding box technique used
	Brain tumor [24]	DeepLab		CT images	DeepLab with conditional random fields produces high accuracy
Lungs	Pulmonary nodules [25]	3D FCN	LIDC dataset	Chest CT images	Increased speed of screening
	Lung segmentation [26]	JU-Net based CNN	JSRT	CXR	____
	Pneumothorax segmentation [27]	FC-DenseNet with SCSE module	PACS	Chest X-ray images	Spatial weighted cross-entropy loss function improves precision at boundaries
	Pneumothorax segmentation [28]	Mask R-CNN	SIIM-ACR	Chest X-ray images	Bounding box regression helps in improving classification
	Pneumothorax segmentation [29]	U-Net and PSPNet	Routine chest CT dataset	Chest CT images	
Liver	Liver and tumor segmentation [30]	Cascaded FCN	DIRCAD dataset	CT and MRI images	Separate set of filters applied at each stage improves segmentation
	Liver segmentation [31]	HED-mask R-CNN	CHAOS challenge	CT and MR images	High segmentation accuracy obtained
	Liver segmentation [32]	FCN	MICCAI SLiver07 dataset	CT images	—
Reproductive system	Prostate [33]	VNet	—	3D MRI	—
Digestive system	Pancreas [34]	Recurrent NN (LSTM)	NIH-CT-82, ufl-mri-79	Abdominal CT and MRI images	RNN performs better than HNN and UNET
Breast	Breast masses [35]	DBN + CRF/ SSVM	DDSM-BCRP, INbreast databases	Mammograms	CRF model is faster than SSVM
Eyes	Retinal blood vessels [36]	U-Net with modifications	DRIVE/STARE	Retinal images	Modification allows precise and faster segmentation of blood vessels
	Retinal blood vessels [37]	U-Net, LadderNet	DRIVE/STARE/ CHASE	Retinal images	—

ADC: Alzheimer Disease Center. MICCAI: Medical Image Computing and Computer Assisted Intervention. MRBrainS: MR brain segmentation. ISBI: IEEE International Symposium on Biomedical Imaging. LIDC: Lung Image Database Consortium. JSRT: Japanese society of radiological technology. PACS: Picture Archiving and Communication System. SIIM-ACR: Society for Imaging Informatics in Medicine-American College of Radiology. DIRCAD: 3D image reconstruction for comparison of algorithm database. CHAOS: combined (CT-MR) healthy abdominal organ segmentation challenge. DDSM: digital database for screening mammography. DRIVE: digital retinal images for vessel extraction. STARE: Structural Analysis of Retinal Dataset. CHASE: Combined Healthy Abdominal Organ Segmentation Challenge.

Table 2 Public Dataset for Medical Image Segmentation

Objects	Dataset	URL
Liver	LiTS	https://competitions.codalab.org/competitions/17094
	Sliver07	http://www.sliver07.org/
	3Dircadb	https://www.ircad.fr/research/3dircadb/
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
	CHAOS	https://chaos.grand-challenge.org
Pancreas	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
	NIH Pancreas	http://academictorrents.com/details/80ecfefcabede760cdbdf63e38986501f7becd49
Colon	COLONOGRAPHY	https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY#dc149b9170f54aa29e88f1119e25ba3e
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Heart	AMRG Cardiac Atlas	http://www.cardiacatlas.org/studies/amrg-cardiac-atlas/
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Lung	LIDC-IDRI	https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#
	VESSEL12	https://vessel12.grand-challenge.org/
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Prostate	PROMISE12	https://promise12.grand-challenge.org/
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Brain	OASIS	http://www.oasis-brains.org/
	BRATS	https://www.med.upenn.edu/sbia/brats2018/registration.html
	ISLES	http://www.isles-challenge.org/
	mTOP	https://www.smir.ch/MTOP/Start2016
	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Kidney	KITS	https://kits19.grand-challenge.org
	CHAOS	https://chaos.grand-challenge.org
Spleen	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
	CHAOS	https://chaos.grand-challenge.org
Hippocampus	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Hepatic Vessel	Medical Segmentation Decathlon (MSD)	http://medicaldecathlon.com/index.html
Skin lesion	ISIC	https://challenge.isic-archive.com/data
STARE	STARE	https://cecas.clemson.edu/ ahoover/stare/
Thyroid	TNSCUI	https://tn-scui2020.grand-challenge.org/

2. Research Challenges

This proposed research aims to solve the problem of medical Image Segmentation using deep learning architectures and methods. The key challenges in medical image segmentation are given in following table 3.

Table 3. Key Challenges in Medical Image Segmentation

Challenge	Description
Variability in Image Acquisition	Medical images can be acquired from different machines with varying protocols, which can lead to variability in the appearance of the images. This can make it difficult to develop segmentation models that generalize well to new images.
Noise and Artifacts	Medical images are often subject to noise and artifacts, which can obscure the underlying anatomy and make it difficult to accurately segment structures of interest.
Anatomical Variability	The human body can exhibit significant anatomical variability between individuals, which can make it challenging to develop segmentation models that can accurately segment structures across different patients.
Class Imbalance	In medical image segmentation, some structures of interest may be relatively small or infrequent in the images, leading to class imbalance. This can make it challenging to develop accurate segmentation models for these structures.
Expert Annotation	Developing accurate ground truth annotations for medical image segmentation can be challenging, as it often requires expert knowledge and can be time-consuming. In addition, inter-observer variability can make it difficult to establish a definitive ground truth for some structures.
Computational Complexity	Medical images can be very large and complex, which can make it computationally expensive to develop and train segmentation models.
Generalization	Segmentation models trained on one dataset may not generalize well to new datasets, particularly if there are differences in the acquisition protocols or patient populations.
Interpretability	It can be challenging to interpret the output of segmentation models, particularly for more complex architectures such as deep neural networks. This can make it difficult to identify errors or to diagnose problems with the segmentation model.

3. Objectives

The rapid recent developments in Medical Imaging processing technology can used to solve the increasing demand for identification and diagnosis of the diseases in early stages. This challenge includes the problem of segmentation of medical images.

- Propose a deep learning algorithm for segmentation of images
- Collect dataset for various medical images (MRI, CT scan, etc.)
- Pre-process theses images.
- Implement proposed deep learning in Python
- Compare results with existing systems
- Make analysis of the experimented results achieved

4. Literature review

The following section describes the literature review on various approaches used for medical image segmentation.

Segmentation

Image segmentation in computer vision is the division of a digital image into several segments (sets of pixels, also known as image objects). The purpose of segmentation is to reduce complexity and/or transform an image's representation into something more relevant and understandable. Since segmenting a huge image into smaller ones makes subsequent processing easier, segmentation is significant in image processing. When these pieces are linked, the complete picture will be covered. The segmentation process may also be influenced by the image's various properties, such as colour or texture. The process of automatically or partially automatically identifying boundaries within a 2D or 3D image is known as medical image segmentation. Medical images are also produced using a wide range of modalities, including X-ray, CT, MRI, microscope, PET, SPECT, endoscopy, OCT, and many others. Details on the deep neural network structures that can be developed for the segmentation of medical images are provided in the next section.

Deep Neural Network Structures

The most crucial method for artificial intelligence is deep learning. A deep learning technique builds an artificial neural network using a number of layers. An input layer, one or more hidden layers, and an output layer make up an artificial neural network (ANN) [22]. A network's input layer receives signals, an output layer decides how to use the input, and between the input and output layers are hidden layers that carry out calculations (shown in Figure 1). Between the input and output layers, a deep neural network has numerous hidden layers. A review of various deep learning neural networks used for picture segmentation is

provided in this section. As illustrated in Figure 2, the various deep neural network topologies typically used for image segmentation can be categorised. However, Figure 3 shows basic layout of deep learning based system.

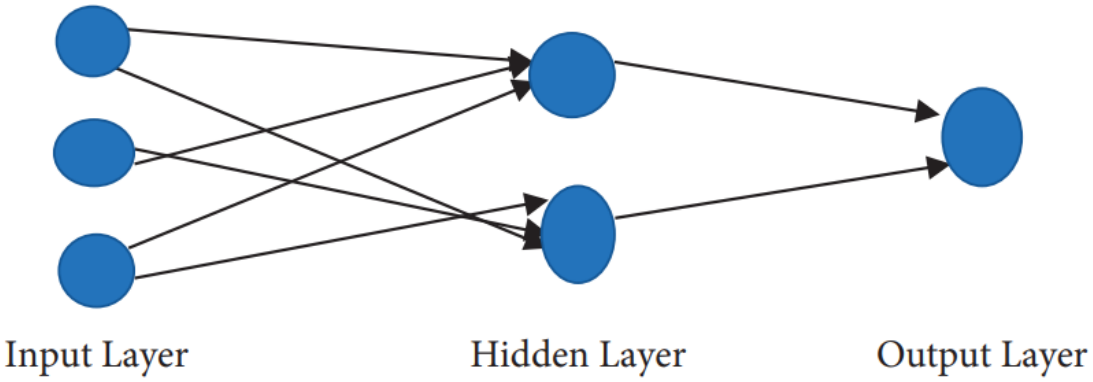

Figure 1. Artificial Neural Network (ANN) Model

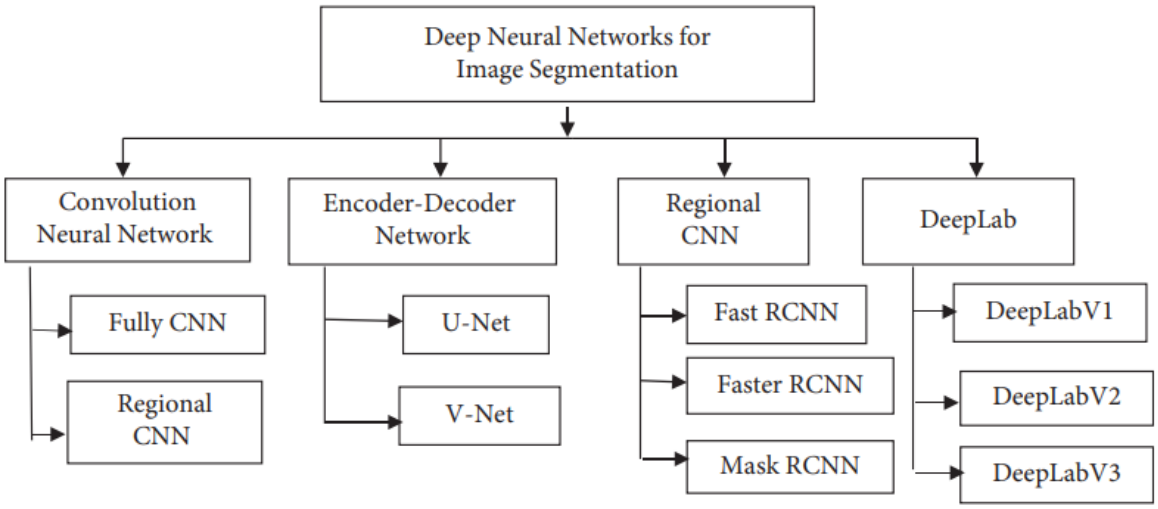

Figure 2. Different Types of Deep Neural Networks for Image Segmentation

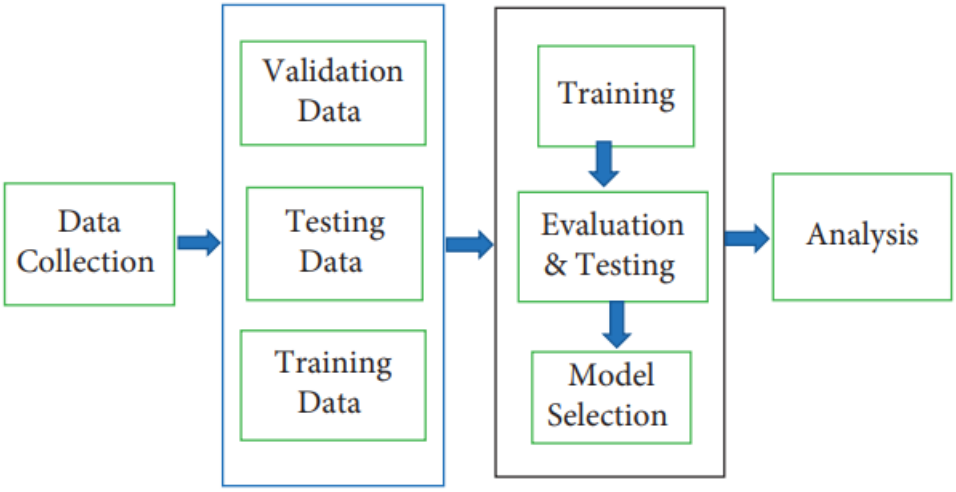

Figure 3 Basic Layout of Deep Learning Based System

Convolutional Neural Network.

Convolutional layer, pooling layer, and fully connected layer are the three basic neural layers that make up a convolutional neural network, or CNN [22, 23]. Each layer plays a distinct role. The convolution layer recognises unique aspects in an image, such as edges or other visual components. With kernels, the convolution layer multiplies a pixel's immediate neighbours to produce an image. For the purpose of creating its feature maps, CNN convolves the given image using several kernels. The spatial (width, height) dimensions of the input data are decreased by the pooling layer for the neural network's subsequent layers. The depth of the data is unaffected. The process is known as subsampling. The computational demands on upcoming layers are reduced as a result of the size reduction. In NN, the fully connected layers perform high-level reasoning. These layers combine the different feature responses from the given input image to produce the final results. In the literature, various CNN models have been reported, including AlexNet [24], GoogleNet [25], VGG [26], Inception[27], SequeezeNet [28], and DenseNet [29]. Each network in this case employs a different number of convolutions and pooling layers, with important process blocks in between. CNN models have primarily been used for classification tasks. SqueezeNet and GoogleNet were used in [20] to classify brain MRI images into three different categories. The following factors limit the performance of CNN segmentation models:

- CNN's fully connected layers cannot handle varying input sizes.
- A convolutional neural network with a fully connected layer cannot be used for object segmentation because the number of objects of interest in the image segmentation task is not fixed, so the length of the output layer cannot be constant.

Fully Convolutional Network.

Only convolutional layers are present in a fully convolutional network (FCN). The various existing CNN architectures can be modified into FCN by converting the last fully connected layer of CNN into a fully convolutional layer. Instead of performing patch-wise predictions, the model designed by [21] can output a spatial segmentation map and perform dense pixel-wise prediction from a full-size input image. The model employs skip connections, which perform upsampling on feature maps from the final layer and fuse them with feature maps from previous layers. Thus, the model generates a detailed segmentation in a single pass. The conventional FCN model, on the other hand, has the following limitations [22]:

- It is slow for real-time inference and does not take into account global context information efficiently. Because of propagation through alternate convolution and pooling layers, the resolution of the feature maps generated at the output is downsampled in FCN.
- This results in FCN predictions with low resolution and fuzziness in object boundaries.

ParseNet [23], an advanced FCN, has also been reported; it uses global average pooling to achieve global context. Models such as conditional random fields and Markov random fields have also been incorporated into DL architecture.

Encoder-Decoder Models.

Encoder-decoder models use a two-stage model to map data points from the input domain to the output domain. The encoder stage compresses the given input, x, to a latent space representation, while the decoder predicts the output from this representation. The following are the various types of encoder-decoder-based models that are commonly used for medical image segmentation:

U-Net.

The U-Net model [24] includes both downsampling and upsampling. To capture context, the downsampling section with FCN-like architecture extracts features using 3 3 convolutions. Deconvolution is used in the upsampling section to reduce the number of computed feature maps. To avoid information loss, the feature maps generated by the downsampling or contracting part are fed as input to the upsampling part. The symmetric upsampling part provides precise localization. The model generates a segmentation map that classifies each pixel in the image.

The U-Net model has the following benefits:

- The U-Net model can perform image segmentation efficiently with a small number of labelled training images.

- The U-Net architecture predicts a fair segmentation map by combining the location information obtained from the downsampling path and the contextual information obtained from the upsampling path.

U-Net models have a few limitations as well, which are as follows:

- The input image size is limited to 572 572 pixels.
- Learning generally slows down in the middle layers of deeper UNET models, causing the network to ignore the layers with abstract features.
- The model's skip connections impose a restrictive fusion scheme, resulting in the accumulation of the same scale feature maps in the encoder and decoder networks.

To address these limitations, the following U-Net architecture variants have been proposed in the literature: U-Net++ [25], Attention U-Net [26], and SD-UNet [27].

VNet.

It is also an FCN-based model used for medical image segmentation [28]. VNet architecture is divided into two parts: the compression network and the decompression network. The compression network is made up of convolution layers with residual functions at each stage. These convolution layers made use of volumetric kernels. The decompression network extracts features and expands the spatial representation of low resolution feature maps. It provides two-channel probabilistic segmentation for both foreground and background regions.

Regional Convolutional Network.

For object detection and segmentation, a regional convolutional network was used. Using a selective search process, the R-CNN architecture presented in [29] generates a region proposal network for bounding boxes. These region proposals are then warped to standard squares and fed into a CNN, which produces a feature vector map as output. The output dense layer is made up of image features that are then fed into a classification algorithm to classify the objects within the region proposal network. The algorithm also predicts offset values in order to improve the precision of the region proposal or bounding box. The following factors limit the use of the basic RCN model:

- It cannot be implemented in real time because training the network for a classification task of 2000 region proposals in a test image takes around 47 seconds.
- The selective search algorithm is predetermined. As a result, learning does not occur at that stage. As a result, unfavourable candidate region proposals may be generated.

To address these shortcomings, various R-CNN variants such as fast R-CNN, faster R-CNN, and mask R-CNN have been proposed in the literature.

Fast R-CNN.

The proposed image overlap regions and the same CNN computations are repeated in R-CNN. [30] describes a fast R-CNN that is fed an input image and a set of object proposals. Convolutional feature maps are then generated by the CNN. Following that, the ROI pooling layer reshapes each object proposal into a fixed-size feature vector. The feature vectors are sent to the model's final fully connected layers. Finally, the computed ROI feature vector is fed to the Softmax layer, which predicts the proposed region's class and offset values [31]. Because of the use of a selective search algorithm, the fast R-CNN is slower.

Faster R-CNN.

The proposed regions in R-CNN and fast R-CNN were created through a slow process of selective search. As a result, in the faster R-CNN architecture described by [32], a single convolutional network was used to perform both region proposals and classification tasks. The model makes use of a region proposal network (RPN), which passes through the sliding window at the top of the CNN feature map. It generates K different potential boundary boxes for each window, with their respective scores representing the position of the object. The precise classification boxes are generated by feeding these bounding boxes to fast R-CNN.

Mask R-CNN.

In [73], He et al. extended faster R-CNN to present Mask R-CNN for example segmentation. The model detects objects in an image and generates a high-quality segmentation mask for each object. It employs the RoI-Align layer to preserve the precise spatial locations of the given image. Using a CNN, the region proposal network (RPN) generated multiple RoIs. The RoI-Align network generates a number of bounding boxes that are warped to fixed dimensions. The previously computed warped features are fed into the fully connected layer to create classification using the softmax layer. The model has three output branches, the first of which computes bounding box coordinates, the second of which determines associated classes, and the third of which evaluates the binary mask for each RoI. The model trains all of the branches at the same time. The regression model is used to improve the bounded boxes. For each RoI, the mask classifier generates a binary mask.

DeepLab Model.

To extract features from an image, the DeepLab model employs a pretrained CNN model ResNet-101/VGG-16 with atrous convolution [34]. Atrous convolutions provide the following advantages:

- It is in charge of the resolution of feature responses in CNNs.

- It transforms an image classification network into a dense feature extractor without the need for additional parameter learning.
- makes use of a conditional random field (CRF) to generate finely segmented output

DeepLab variants such as DeepLabv1, DeepLabv2, DeepLabv3, and DeepLabv3+ have been proposed in the literature.

The input image is passed through a deep CNN layer with one or two atrous convolution layers in DeepLabv1 [37]. This produces a rough feature map. The feature map is then upsampled to the original image size via bilinear interpolation. To obtain the final segmented image, the interpolated data is used to fully connect the conditional random field.

Multiple atrous convolutions are applied to the input feature map at different dilation rates in the DeepLabv2 model. The outputs are combined. Atrous spatial pyramid pooling (ASPP) divides objects into scales. The ResNet model employed atrous convolution with varying rates of dilation. Atrous convolution can capture information from a large effective field with a small number of parameters and computational complexity.

DeepLabv3 [20] is a DeepLabv2 extension that adds image level features to the atrous spatial pyramid pooling (ASPP) module. It also makes use of batch normalisation to make training the network easier. DeepLabv3+ model combines DeepLabv3's ASPP module with encoder and decoder structures. For feature extraction, the model employs the Xception model. To compute faster, the model also used atrous and depth-wise separable convolution. The decoder section combines low- and high-level features that correspond to structural and semantic information.

DeepLabv3+ [76] is made up of two modules: encoding and decoding. The encoding path uses atrous convolution and backbone networks such as MobileNetv2, PNASNet, ResNet, and Xception to extract the required information from the input image. Using information from the encoder path, the decoding path rebuilds the output with relevant dimensions.

Comparison of Different Deep Learning-Based Segmentation Methods.

The different deep neural networks discussed in the above sections are employed for different applications. Each model has its own advantages and limitations. Table 3 gives a brief comparison between different deep learning-based image segmentation algorithms.

Deep learning algorithm	Algorithm description	Advantages	Limitations
CNN	It consists of three main neural layers, which are convolutional layers, pooling layers, and fully connected layers	(a) It is simple (b) It involves feeding segments of an image as input to the network, which labels the pixels	(a) It cannot manage different input sizes (b) Fixed size of output layer causes difficulty in segmentation task
FCN	All fully connected layers of CNN are replaced with the fully convolutional layers	The model outputs a spatial segmentation map instead of classification scores	It is hard to train a FCN model to get good performance
U-Net	It combines the location information obtained from the downsampling path and the contextual information obtained from upsampling path to predict segmentation map	It can perform efficient segmentation of images using limited number of labelled training images	(a) Input image size is limited to 572×572. (b) The skip connections of the model impose a restrictive fusion scheme causing accumulation of the same scale feature maps of the network
VNet	It performs convolutions on each stage using volumetric kernels of size $5 \times 5 \times 5$	It can be applied to 3D data for segmentation	
R-CNN	It uses selective search algorithm to extract 2000 regions from the image called region proposals	(a) It predicts the presence of an object within the region proposals (b) It also predicts four offset values to increase the precision of the bounding box	(a) Huge amount of time is needed to train network to classify 2000 region proposals per image (b) It cannot be implemented in real time (c) Selective search algorithm is a fixed algorithm
Fast R-CNN	It uses selective search algorithm which takes the whole image and region proposals as input in its CNN architecture in one forward propagation	It improves mean average precision (mAP) as compared to R-CNN	There is high computation time due to selective search region proposal generation algorithm
Faster R-CNN	It uses region proposal network	It generates the bounding boxes of different shapes and sizes	There is lower computation time
Mask R-CNN	It gives three outputs for each object in the image: its class, bounding box coordinates, and object mask	a) It is simple and flexible approach b) It is current state-of-the-art technique for image segmentation task	There is high training time
DeepLabv1	a) It uses atrous convolution to extract the features from an image b) It also uses conditional random field (CRF) to capture fine details	a) There is high speed due to atrous convolution b) Localization of object boundaries improved by combining DCNNs and probabilistic graphical models	Use of CRFs makes algorithm slow
DeepLabv2	It uses *atrous spatial pyramid pooling (ASPP)* and applies multiple atrous convolutions with different sampling rates to the input feature map and fuses them together	Atrous spatial pyramid pooling (ASPP) robustly segments objects at multiple scales	There are challenges in capturing fine object boundaries
DeepLabv3	It uses atrous separable convolution to capture sharper object boundaries	It can segment sharper targets	It still needs more refinement for object boundaries
DeepLabv3+	It extends DeepLabv3 by adding a decoder module to refine the segmentation results along the object boundaries	There is better segmentation performance as compared to deepLabv3	It is a large model with number of parameters to train. So, while training on higher resolution images and batch sizes, it needs large GPU memory.

Table 3 Comparison of various segmentation techniques

5. Proposed Work

The proposed study focuses on identifying fake information. The study suggests a hybrid model which is based on Content-based factors, Title based factors, and author-based factors. In the model, a major focus is on content, catchy titles, writing styles, the legacy or credibility of the author, and its behavior. All these factors are applied in supervised machine learning algorithms on the dataset. The dataset used for the study was downloaded from the online available dataset and is freely available.

6. Methodology

The proposed study focuses on medical image segmentation. The study suggests a deep learning model which is based on U-net architecture. The dataset used for the study is downloaded from the online available website and is freely available.

Dataset:

In this research work, datasets play a major role in medical image segmentation. Dataset is a collection of records or information which is from different fields. Dataset is in image format and downloaded from the online available free sources. It may contain MRI and CT scan images, etc.

The following figure depicts proposed model for medical image segmentation.

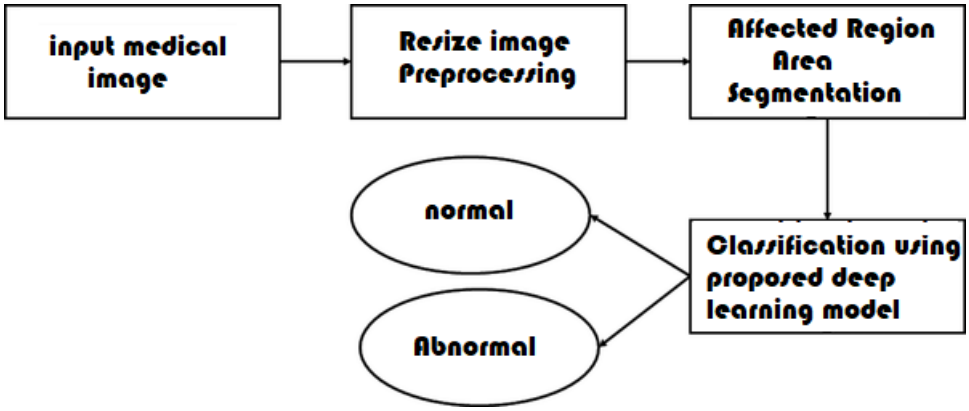

Figure 4 Proposed model for early diagnosis of diseases

H/W and S/W Requirements:-

This model is implemented in python and anaconda. For implementation purposes, different libraries of python are used. The libraries which are used for implementation purposes are numpy, scipy, Scikit-Learn, TensorFlow, Keras, Pandas, Matplotlib, OpenCV, etc.

Evaluation Techniques

The evaluation technique helps you to assess your model's performance, monitor, and control. We aim to create a model which gives high accuracy on out-of-sample data. Evaluation metrics are tied to machine learning tasks. Using different metrics for the tasks of classification describes the performance of a model on a set of test data. Correct and incorrect predictions are grouped by class.

a. Accuracy: the percentage of correct predictions for the test data. the ratio between the number of correct predictions to the total number of predictions.

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

i.e. Accuracy= number of correct predictions / Total number of predictions

b. Precision: To identify the correctness of classification. The ratio between the number of *Positive* samples correctly classified to the total number of samples classified as *Positive* (either correctly or incorrectly). Higher precision means better the ability of the model to correctly classify the positive class.

$$Precision = \frac{True_{positive}}{True_{positive} + False_{positive}}$$

i.e. Precision = Number of correct positive results / Total number of positive results

c. Recall: Recall tells us the number of positive cases correctly identified out of the total number of positive cases. A high recall value indicates that a lot of fake cases are identified out of the total number of fakes.

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}}$$

Recall = Number of correct positives/number of all positives

d. F1- Score: weighted average of precision and recall. Precision is used to save the company's costs. The recall is used to ensure the machinery is stable and not a threat to human lives[22]. It is a measure of a test's accuracy

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}$$

In machine learning, these metrics are commonly used to evaluate classifier performance from different angles. Accuracy measures the similarity between disease predictions and actual disease information.

7. Result Analysis

Standard metrics measure the overall performance of each algorithm in the classification process. A matrix illustrating the results of each classifier on the test data is shown by the number of correct and incorrect predictions for each class. Each model's performance was is evaluated in detail to provide a comprehensive result analysis. This study aims to compare all algorithms in terms of their overall performance.

8. Conclusion

Medical imaging is an important part of today's healthcare system because it allows for non-invasive diagnostic procedures. It entails creating visual and functional representations of the

human body's interior and organs for clinical analysis. X-ray-based methods, such as conventional X-ray, computed tomography (CT), and mammography; molecular imaging, magnetic resonance imaging (MRI), and ultrasound (US) imaging are among the various types. Aside from these medical imaging techniques, clinical images are increasingly being used to diagnose a variety of conditions, particularly those related to the skin.

Medical imaging is divided into two parts: image formation and reconstruction and image processing and analysis. Image formation refers to the process of creating two-dimensional (2D) images of three-dimensional (3D) objects, whereas reconstruction uses a set of iterative algorithms to create 2D and 3D images from an object's projection data. Image processing, on the other hand, entails using algorithms to improve image properties such as noise removal, whereas image analysis extracts quantitative information or a set of features from an image for object identification or classification.

Because of advances in technology, it is now possible to acquire large amounts of high-resolution images at very low cost. As a result, the development of biomedical image processing algorithms has advanced significantly. As a result, automated image analysis or evaluation algorithms for extracting useful information have been developed. The first step in such automated analysis is segmentation, which divides the image into visually distinct regions with semantic meaning for the given problem. Each of these regions has consistent characteristics in terms of grey level, texture, or colour. Further analysis, such as determining texture homogeneity levels or layer thickness, requires clear segmentation and distinguishable regions. At times, the image may contain multiple objects of the same class. In addition, the segmentation process may segregate regions containing objects of the same class while ignoring objects of other classes, which is known as instance segmentation, as opposed to semantic segmentation, which segregates objects of the same class but not objects of different classes.

All image segmentation techniques can be classified into three types: 1) Manual segmentation (MS), 2) semi-automatic segmentation, and 3) fully automatic segmentation techniques are available. The essential capability of deep learning approaches to obtain hierarchical feature representations directly from images, which eliminates the need for handcrafted features, is at the heart of these advancements. Deep learning is quickly becoming the state-of-the-art for medical image processing, resulting in performance improvements in a variety of clinical applications. The proposed research seeks to achieve fully automatic segmentation. To be specific, the research aims to automate image segmentation in the field of medical imaging using deep learning approaches. Current advances in machine learning, particularly deep

learning, are proving useful in identifying and quantifying patterns in medical images. We also intend to develop image segmentation implementations for a variety of medical applications.

9. References

[1].	A. A. Albishri, S. J. H. Shah, and Y. Lee, "Cu-net: Cascaded u-net model for automated liver and lesion segmentation and summarization," Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM), pp. 1416–1423, 2019.
[2].	A. A. Kalinin, V. I. Iglovikov, A. Rakhlin, and A. A. Shvets, "Medical image segmentation using deep neural networks with pre-trained encoders." Springer, 2020, pp. 39–52.
[3].	A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
[4].	A. Boutillon, B. Borotikar, V. Burdin, and P.-H. Conze, "Combining shape priors with conditional adversarial networks for improved scapula segmentation in mr images," Proc. IEEE Int. Symp. Biomed. Imag. (ISBI), pp. 1164–1167, 2020.
[5].	A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. Newby, R. Dharmakumar, and S. A. Tsaftaris, "Factorised spatial representation learning: Application in semi-supervised myocardial 22 segmentation," Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI), pp. 490–498, 2018.
[6].	A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," Int. Workshop Simul. Synth. Med. Imag, pp. 3–13, 2017.
[7].	A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly et al. , "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
[8].	A. E. Kavur, M. A. Selver, O. Dicle, M. Barıs, and N. S. Gezer, "Chaoscombined (ct-mr) healthy abdominal organ segmentation challenge data," 2019.
[9].	A. E. Kavur, N. S. Gezer, M. Barıs¸, P.-H. Conze, V. Groza, D. D. Pham, S. Chatterjee, P. Ernst, S. Ozkan, B. Baydar ¨ et al., "Chaos challenge–combined (ct-mr) healthy abdominal organ segmentation," arXiv preprint arXiv:2001.06535, 2020.
[10].	A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
[11].	A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE

	Transactions on Medical imaging, vol. 19, no. 3, pp. 203–210, 2000.
[12].	A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.
[13].	A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," arXiv preprint arXiv:1902.09063, 2019.
[14].	A. Suinesiaputra, P. Medrano-Gracia, B. R. Cowan, and A. A. Young, "Big heart data: advancing health informatics through data sharing in cardiovascular imaging," IEEE J. Biomed. Health Inform., vol. 19, no. 4, pp. 1283–1290, 2014.
[15].	A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets," IEEE Trans. Med. Imaging, vol. 22, no. 2, pp. 137–154, 2003.
[16].	A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8543–8553, 2019.
[17].	B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest et al., "The multimodal brain tumor image segmentation benchmark (brats)," IEEE Trans. Med. Image., vol. 34, no. 10, pp. 1993–2024, 2014.
[18].	B. Yang, H. Pan, J. Yu, K. Han, and Y. Wang, "Classification of medical images with synergic graph convolutional networks," 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), pp. 253–258, 2019.
[19].	B. Zhang, J. Xiao, J. Jiao, Y. Wei, and Y. Zhao, "Affinity attention graph neural network for weakly supervised semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[20].	C. Baur, S. Albarqouni, and N. Navab, "Semi-supervised deep learning for fully convolutional networks," Proc. Int. Conf. Med. Image Comput. Comput-Assist. Intervent., pp. 311–319, 2017.
[21].	C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: a review," Frontiers in Cardiovascular Medicine, vol. 7, p. 25, 2020.
[22].	C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 865–872, 2019.
[23].	C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations."

	Springer, 2017, pp. 240–248.
[24].	C. Huang, H. Han, Q. Yao, S. Zhu, and S. K. Zhou, "3d u net: A 3d universal u-net for multi-domain medical image segmentation," Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI), pp. 291–299, 2019.
[25].	C. Kaul, S. Manandhar, and N. Pears, "Focusnet: An attention-based fully convolutional network for medical image segmentation," Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI), pp. 455–458, 2019. 21
[26].	C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, and D. D. Feng, "A likelihood and local constraint level set model for liver tumor segmentation from ct volumes," IEEE Trans. Biomed. Eng., vol. 60, no. 10, pp. 2967–2977, 2013.
[27].	C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei, "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation," Proc. the IEEE Conf. Comput. Vis. Pattern Recognitit. (CVPR), pp. 82–92, 2019.
[28].	C. Rother, V. Kolmogorov, and A. Blake, "" grabcut" interactive foreground extraction using iterated graph cuts," ACM Trans. Graph., vol. 23, no. 3, pp. 309–314, 2004.
[29].	C. Rupprecht, I. Laina, N. Navab, G. D. Hager, and F. Tombari, "Guide me: Interacting with deep networks," Proc. Int. Conf. Med. Image Comput. Comput-Assist. Intervent, pp. 8551–8561, 2018.
[30].	C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," Proc. the IEEE Conf. Comput. Vis. Pattern Recognitit.(CVPR), pp. 2818–2826, 2016.
[31].	C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Proc. the IEEE Conf. Comput. Vis. Pattern Recognitit.(CVPR), pp. 1–9, 2015.
[32].	C. Wang, Y. He, Y. Liu, Z. He, R. He, and Z. Sun, "Sclerasegnet: an improved u-net model with attention for accurate sclera segmentation," Proc. IAPR Int. Conf. Biometrics, pp. 1–8, 2019.
[33].	C. Zhao, A. Carass, J. Lee, Y. He, and J. L. Prince, "Whole brain segmentation and labeling from ct using synthetic mr images," Int. Workshop Mach. Learn. Med. Imag, pp. 291–298, 2017.
[34].	C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeplysupervised nets," Artif. Intell. Statistics, pp. 562–570, 2015.
[35].	D. Jin, Z. Xu, Y. Tang, A. P. Harrison, and D. J. Mollura, "Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation," Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI), pp. 732–740, 2018.

[36].	D. Karimi and S. E. Salcudean, “Reducing the hausdorff distance in medical image segmentation with convolutional neural networks,” IEEE Trans. Med. Imag., vol. 39, no. 2, pp. 499–513, 2019.
[37].	D. Mahapatra, B. Bozorgtabar, J.-P. Thiran, and M. Reyes, “Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network,” Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI), pp. 580–588, 2018.